UNIVERSITY OF MANNHEIM

Testing Sample-Based Accounts of Probability Judgements Using a Ranking Task

Xiaotong Liu, July 2021

SMiP IOPS Conference

STATISTICAL MODELING in PSYCHOLOGY
FREIBURG HEIDELBERG LANDAU MANNHEIM TÜBINGEN

Introduction

- People's explicit probability judgements often appear to be probabilistically incoherent (i.e., Linda problem)

- Biases in probability judgements can arise from rational reasoning processes based on mental samples

- Sample-based probability judgement models assume that...

  ...samples generated from probability distributions in mind
  ...each sample takes time to generate
  ...infinite number of samples result in unbiased judgements

One sample-based model: PT+N model

Core assumptions in Probability Theory Plus Noise (PT+N) model (Costello & Watts, 2014):

- DM holds coherent sets of beliefs such that $P(a) + P(\neg a) = 1$
- When asked to explicitly estimate $P(A)$...
  ...DM requests N samples, according to **a fixed belief**, $P(a)$
  ...DM counts number of critical samples i that indicate A.
  ...sample has some chance d of being mistaken for its complement

  $\Rightarrow$ estimation based on relative frequencies of number of critical samples

One sample-based model: PT+N model

From these assumptions,

- Expected value of individual estimates (Costello & Watts, 2014, Eq. 3)
  $P_E(A) = (1 - 2d)P(a) + d$

- Estimate of number of critical samples (Howe & Costello, 2020, Eq. 4)
  $P(i|(N, p) = Bin(N, p) = \binom{N}{i} p^i (1 - p)^{N-i}$
  $p = P_E(A)$
  N: total sample size; i: the number of critical events

The limitation

- PT+N account only tested in estimation tasks
  - Mono-operation bias
  - Sampling parameters not identifiable without additional assumptions

- A novel ranking task is proposed
  - Ranking task
  - PT+N rank model
  - Derive qualitative predictions
  - Test qualitative predictions experimentally

The ranking Task

Please rank the four events below according to their likelihoods.

Assign **rank 1 to the event that you think is most likely to occur**, rank 2 to the event that you think is second most likely to occur, and so forth. Thus the event that you think is least likely to occur should receive rank 4. **Please specify the rank by selecting the corresponding number on the right.**

	1	2	3	4
On a random day in England, it will **not** be rainy.	○	○	○	○
On a random day in England, it will be cold.	○	○	○	○
On a random day in England, it will be rainy.	○	○	○	○
On a random day in England, it will **not** be cold.	○	○	○	○

Normative rules in ranking task

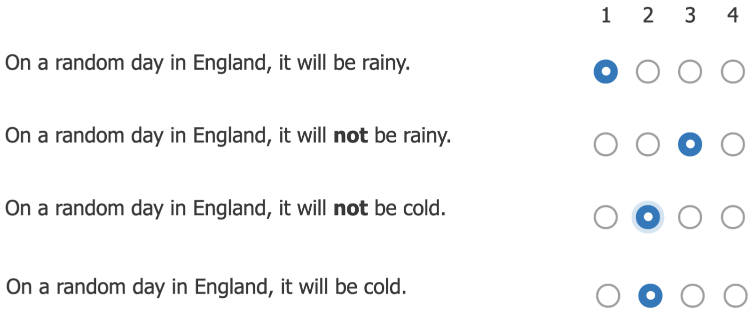
- The complement rule requires $P(a) + P(\neg a) = 1$
  - For logically possible rankings, $\text{Rank}(A) + \text{Rank}(\neg A) = \text{Rank}(B) + \text{Rank}(\neg B)$

- 75 possible ranks over four events
  - **17 of them are logically possible**
  - Rankings with vs. without ties
  - possibility to categorise logically impossible rankings

Normative rules in ranking task

Logically possible rankings. . .

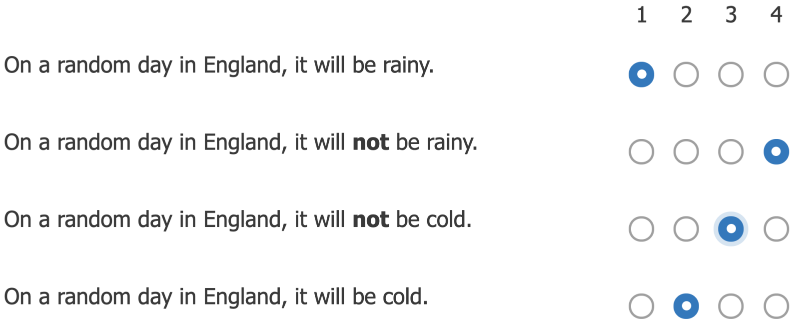
- With ties
  i.e., $\text{Rank}(A) + \text{Rank}(\neg A) = 4$; $\text{Rank}(B) + \text{Rank}(\neg B) = 4$



Normative rules in ranking task

Logically possible rankings. . .

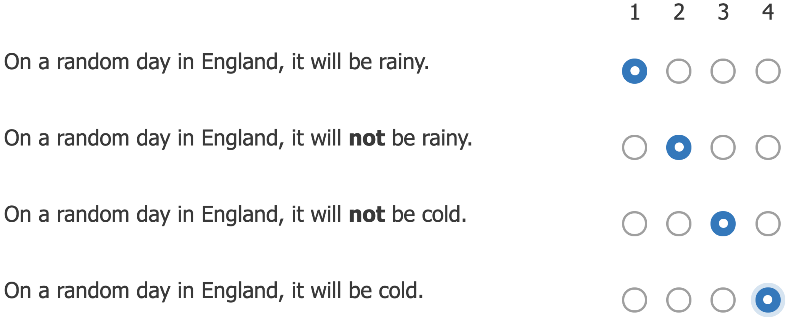
- Without ties
  i.e., $\text{Rank}(A) + \text{Rank}(\neg A) = 5$; $\text{Rank}(B) + \text{Rank}(\neg B) = 5$



Normative rules in ranking task

Categorise rankings without ties. . .

- class one illogical rankings:
  i.e., A & $\neg A$ simultaneously ranked above B & $\neg B$



Normative rules in ranking task

Categorise rankings without ties...

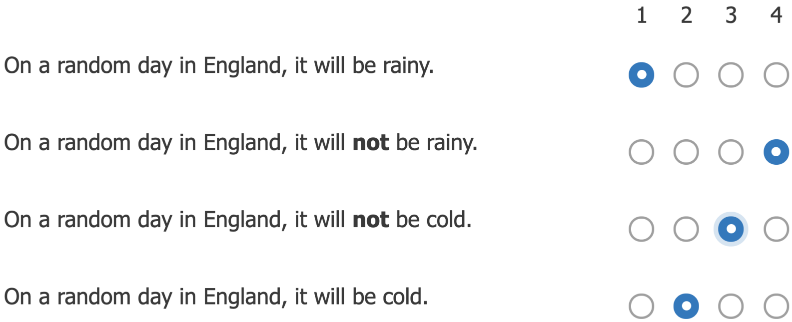
- class two illogical rankings:
  i.e., $A > B$ & $\neg A > \neg B$ simultaneously hold in one response

	1	2	3	4
On a random day in England, it will be rainy.	●	○	○	○
On a random day in England, it will **not** be rainy.	○	○	●	○
On a random day in England, it will **not** be cold.	○	○	○	●
On a random day in England, it will be cold.	○	●	○	○

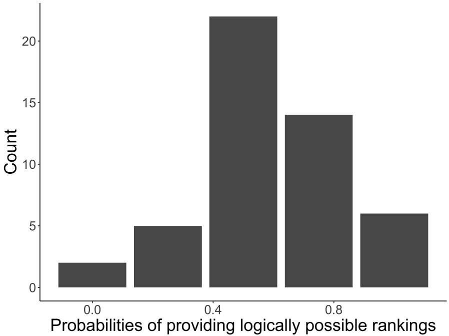
Normative rules in ranking task

Categorise rankings without ties. . .

- logically possible rankings



A new bias?



49 participants were asked to rank $\{\text{event } A, \text{ event} \neg A, \text{ event} B, \text{ event} \neg B\}$.

On average, around two out of the four rankings that participants provided were logically possible (mean $= 0.59$, SD $= 0.24$).

A new bias?

Is it a real bias, or merely the consequence of mental sampling processes?

A not so clear answer: PT+N can produce logically impossible rankings. But we need to test its sufficiency in explaining rankings.

PT+N rank model

What does PT+N model imply the rank-based probability judgements?

PT+N rank model:

- model assumptions:
  - DM holds coherent sets of beliefs
  - When asked to rank a set of events (i.e., $\{A, \neg A, B, \neg B\}$)...
    ...DM implements direct sampling for each event in set
    ...**for events in same set, one uses same sample size**
    ...**derives ranks by comparing number of critical samples**
  - sample has some chance d of being mistaken for its complement

PT+N rank model

Analytic formula for rankings without any ties:

$$P(A > B > C > D) = P(C_A > C_B > C_C > C_D)$$

$$= \sum_{i_d=0}^{N-3} f(i_d, N, d) \sum_{i_c=(i_d+1)}^{N-2} f(i_c, N, c) \sum_{i_b=(i_c+1)}^{N-1} f(i_b, N, b) \sum_{i_a=(i_b+1)}^{N} f(i_a, N, a)$$

$$f(i, N, p) = \binom{i}{N} p^i (1-p)^{N-i}$$

N: total sample size; i: the number of critical event; p: the expected value of P(A)

Assume N = 4:

i_A	i_B	i_C	i_D
3	2	1	0
4	2	1	0
4	3	1	0
4	3	2	0
4	3	2	1

- Substitute A, B, C, D with A, $\neg A$, B, $\neg B$ in 24 linear orderings

$\{A, \neg A, B, \neg B\}$
$\{A, \neg A, \neg B, B\}$
$\{A, B, \neg A, \neg B\}$
......

Predictions for rankings

Analytic formula for ranking with ties:

$$P(A > B > C = D)$$
$$= \sum_{i_{c,d}=0}^{N-2} f(i_{c,d}, N, d) f(i_{c,d}, N, c) \sum_{i_b=(i_{c,d}+1)}^{N-1} f(i_b, N, b) \sum_{i_a=(i_b+1)}^{N} f(i_a, N, a)$$
$$P(A > B = C > D) = \ldots$$
$$P(A = B > C > D) = \ldots$$

$$P(A = B > C = D)$$
$$= \sum_{i_{c,d}=0}^{N-1} f(i_{c,d}, d) f(i_{c,d}, n, c) \sum_{i_{a,b}=(i_{c,d}+1)}^{N} f(i_{a,b}, b) f(i_{a,b}, a)$$

$$P(A = B = C > D)$$
$$= \sum_{i_d=0}^{N-1} f(i_d, N, d) \sum_{i_{a,b,c}=(i_d+1)}^{N} f(i_{a,b,c}, N, c) f(i_{a,b,c}, N, b) f(i_{a,b,c}, N, a)$$
$$P(A > B = C = D) = \ldots$$

$$P(A = B = C = D)$$
$$= \sum_{i_{a,b,c,d}=0}^{N} f(i_{a,b,c,d}, N, d) f(i_{a,b,c,d}, N, c) f(i_{a,b,c,d}, N, b) f(i_{a,b,c,d}, N, a)$$

An empirical issue

Will tied rankings cause troubles on model testing?

We hypothesise that if ties are allowed, lazy participants will overuse ties.

The solution

- add additional assumptions:
  - ...
  - ...
  - ...
  - **When tied rankings not allowed, participants will produce a random linear ordering among tied events but still follow the ordering among the untied events**

The solution

	1	2	3	4
On a random day in England, it will be rainy.	●	○	○	○
On a random day in England, it will **not** be rainy.	○	○	●	○
On a random day in England, it will **not** be cold.	○	●	○	○
On a random day in England, it will be cold.	○	●	○	○

1	2	3	4
●	○	○	○
○	○	○	●
○	○	●	○
○	●	○	○

1	2	3	4
●	○	○	○
○	○	○	●
○	●	○	○
○	○	●	○

Derive qualitative predictions

- How do we derive the qualitative predictions?
  - Four parameters enter the model, N, d, A(or $\neg A$), B (or $\neg B\}$).
  - There categories of the responses
    - class one illogical rankings
    - class two illogical rankings
    - logically possible rankings

Derive qualitative predictions

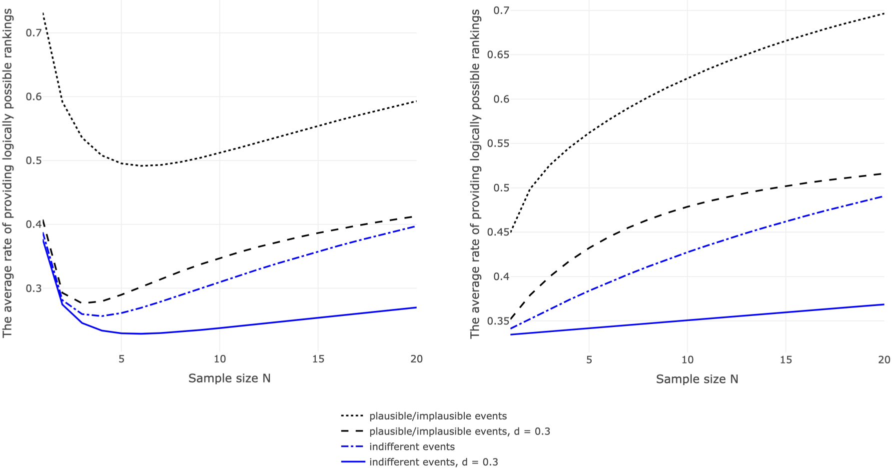
Simulation

- Inputs: different parameter values
  - Inherent beliefs (i.e., $P(a)$)
    - beliefs about plausible/implausible events, drawn from Beta(1, 10)
    - beliefs about indifferent events, drawn from Beta(10, 10)
  - reading error d
    - $d = 0$ or 0.3
  - sample size N
    - range from 1 to 20 in a step of 1
- Repeat 100000 times
- Results
  - simulate ranking data for two types of event sets
    - plausible/implausible vs. indifferent event sets
  - calculate prob. of providing three categories of rankings

Derive qualitative predictions

Average rate of producing logically possible rankings

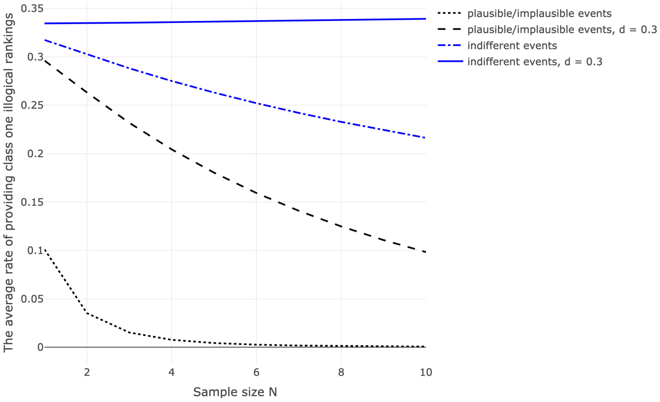
left: allow ties vs. right: not allow ties



Derive qualitative predictions

Average rate of producing class one illogical rankings

ties not allowed



Testing the qualitative predictions experimentally

Design

- manipulate inherent beliefs within subjects
  - ask participants to rank two types of events
    plausible/implausible vs. indifferent event set
- manipulate whether or not ties are allowed between subjects

Hypothesis

If we compare plausible/implausible event sets to indifferent event sets,

- Average rate of providing class one illogical rankings will be lower
- Average rate of providing logically possible rankingswill be higher

Summary

- Ranking task to investigate coherence of beliefs in a novel way
  - Biases in rank-based probability judgements
    - can be explained by existing sample-based models

- Need to understand and test the consequences of sampling
  - Ranking tasks allow us to do so
    - testing qualitative predictions of rankings

Discussion

- Future research ideas
  - Model is identifiable
    $\Rightarrow$ parameters can be estimated on individual level
  - Manipulate sample size N by...
    ...manipulating response time
    ...manipulating the structure of sentences
    - A randomly selected resident is living in southern Germany.
    - There are more people living in southern Germany than in northern Germany.

Thank you!

This research is a joint work by

Xiaotong Liu (University of Mannheim),

Henrik Singmann (University of College London),

and Arndt Bröder (University of Mannheim).

Thank you!

STATISTICAL MODELING in PSYCHOLOGY

FREIBURG HEIDELBERG LANDAU MANNHEIM TÜBINGEN

This research was funded by the Deutsche Forschungsgemeinschaft (DFG), grant 2277, Research Training Group “Statistical Modeling in Psychology" (SMiP)